

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平11-259485

(43) 公開日 平成11年(1999) 9月24日

(51) Int. Cl.⁵
G 0 6 F 17/30

識別記号

F I
G 0 6 F 15/403

3 5 0 C
3 8 0 E

審査請求 未請求 請求項の数6 O L (全 18 頁)

(21) 出願番号 特願平10-54812

(22) 出願日 平成10年(1998) 3月6日

(71) 出願人 000005223

富士通株式会社

神奈川県川崎市中原区上小田中4丁目1番
1号

(72) 発明者 小川 知也

神奈川県川崎市中原区上小田中4丁目1番
1号 富士通株式会社内

(74) 代理人 弁理士 今村 辰夫 (外1名)

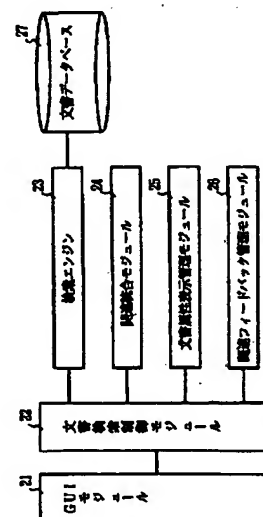
(54) 【発明の名称】 文書検索装置及び記録媒体

(57) 【要約】

【課題】本発明は文書検索装置に関し、文書検索時に得られる文書データからの情報を活用して、効率的な検索結果提示や操作手段の提供を可能とする。

【解決手段】文書データベース27と、文書データベース27の検索処理を行う検索エンジン23と、情報の視覚的操作のためのインターフェース制御を行うGUIモジュール21と、検索結果情報の関連統合処理を行う関連統合モジュール24と、文書属性の表示管理を行う文書属性表示管理モジュール25と、検索結果情報の評価情報を管理し、関連フィードバックが行われた際、その文書の表示情報を制御する関連フィードバック管理モジュール26とを備え、文書検索での検索結果提示の際に、検索結果の文書間の関連度情報に基づき、関連度の高い文書をまとめて表示するように構成した。

文書検索モジュールの説明図



【特許請求の範囲】

【請求項1】文書データを格納した文書データベースと、前記文書データベースの文書検索を行う検索エンジンと、情報の視覚的操作のためのインターフェース制御を行うグラフィカルユーザインターフェースモジュールと、前記検索エンジンによる検索結果の情報に対して関連情報の統合処理を行う関連統合モジュールと、文書属性の表示管理を行う文書属性表示管理モジュールと、検索結果情報の評価情報を管理し、関連フィードバックが行われた際その文書の表示情報を制御する関連フィードバック管理モジュールと、

を備えた文書検索装置であって、
文書検索での検索結果提示の際に、検索結果の文書間の関連度情報に基づき、関連度の高い文書の情報をまとめて表示させる検索結果情報提示手段を備えていることを特徴とした文書検索装置。

【請求項2】前記検索結果情報提示手段は、関連度の高い文書の情報をまとめて表示させる際、多段階に階層的に表示させる機能を備えていることを特徴とした請求項1記載の文書検索装置。

【請求項3】前記検索結果情報提示手段は、検索結果の文書間の関連度情報に基づき、上位レベルの文書のみを表示や下位レベルの展開を含むアウトライン操作を行う機能を備えていることを特徴とした請求項1記載の文書検索装置。

【請求項4】前記検索結果情報提示手段は、検索結果提示の際に、同一グループに属する複数文書に対し、関連フィードバックの対象とする／しないの指定を一度に行える手段を提供する機能を備えていることを特徴とした請求項1記載の文書検索装置。

【請求項5】前記検索結果情報提示手段は、検索結果提示の際に、関連度判定の閾値を、アイテムの密集度に応じて動的に変更する機能を備えていることを特徴とした請求項1記載の文書検索装置。

【請求項6】文書データを格納した文書データベースと、前記文書データベースの文書検索処理を行う検索エンジンと、情報の視覚的操作のためのインターフェース制御を行うグラフィカルユーザインターフェースモジュールと、前記検索エンジンによる検索結果の情報に対して関連情報の統合処理を行う関連統合モジュールと、文書属性の表示管理を行う文書属性表示管理モジュールと、検索結果情報の評価情報を管理し、関連フィードバックが行われた際、その文書の表示情報を制御する関連フィードバック管理モジュールとを備えた文書検索装置に、

文書検索での検索結果提示の際に、検索結果の文書間の関連度情報に基づき、関連度の高い文書をまとめて表示させる手順を実行させるためのプログラムを記録したコンピュータ読取可能な記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、文書データベースから検索した検索結果の情報を、ディスプレイ装置の画面上に表示する機能を備えた文書検索装置及び記録媒体に関する。

【0002】近年、インターネットの普及や全文データベースの増加に伴い、我々の身の回りの情報は飛躍的に増加している。しかし、検索精度や検索システムの使い勝手に関し、現状の大規模な商用のオンラインデータベースや全文検索システムは、必ずしも十分とはいえない。必要な情報が効率良く入手できる情報検索技術の開発が望まれている。

【0003】

【従来の技術】以下、従来例を説明する。図16は従来の検索結果情報表示例を示した図である。従来、文書検索における検索結果の提示では、検索結果を検索キーとの文書間の関連度に基づき、一次的に表示するランキング表示が行われている程度であった。このような文書検索における検索結果情報の表示例として、例えば、図16に示したような例がある。なお、この場合の「検索結果の提示」は、例えば、表示であるが、表示に限らず、他の出力手段（印刷出力等）により利用者に提示できれば良い。

【0004】この例は、数百件の文書（論文）からの検索を行った場合の例であり、「概念学習」を検索キーとして検索を行うと、図16に示したような検索結果情報が提示される。この例では、検索結果の提示として、順位、関連度、文書ID、及び文書タイトルについての各情報が提示されている。

【0005】前記関連度は、通常のベクトル空間モデルや特異値分解などの統計処理に基づく意味空間モデルにおける検索キーベクトルと、各文書ベクトルとのCOS距離などで求める。この場合、前記COS距離は、前記2つのベクトルの成す角度を θ とした場合、 $COS = COS(\theta)$ などで求めることができる。

【0006】前記の表示例において、順位=1は、関連度=0.84、文書ID=92-89-03であり、文書タイトルは、「名詞述語文からの概念学習システムー比喩的理解と学習ー」となっている。順位=2は、関連度=0.78、文書ID=90-79-04であり、文書タイトルは、「名詞述語文における意味概念の学習」となっている。

【0007】順位=3は、関連度=0.74、文書ID=91-83-19であり、文書タイトルは、「自然言語概念学習に関する考察」となっている。順位=4は、関連度=0.72、文書ID=93-93-06であり、文書タイトルは、「名詞述語文からの概念学習システムー継承と経路情報の処理ー」となっている。以下、同様に、一次的検索結果が提示されている。

【0008】

【発明が解決しようとする課題】前記のような従来のものにおいては、次のような課題があった。すなわち、従来の文書検索における検索結果提示では、検索結果情報を検索キーとの関連度に基づき、一次的に表示するランキング表示が行われている程度であり、検索時に得られる文書データからの情報を十分に活用した検索結果の表示や、操作手段の提示は行われていなかった。従って、必要な情報が効率良く入手できなかった。

【0009】本発明は、このような従来の課題を解決し、文書検索時に得られる文書データからの情報を有効に活用することで、効率的な検索結果情報の提示や操作手段の提供を可能にすることを目的とする。

【0010】

【課題を解決するための手段】本発明は前記の目的を達成するため、次のように構成した。

(1)：文書データを格納した文書データベースと、前記文書データベースの文書検索を行う検索エンジンと、情報の視覚的操作のためのインターフェース制御を行うグラフィカルユーザインターフェースモジュールと、前記検索エンジンによる検索結果の情報に対して関連情報の統合処理を行う関連統合モジュールと、文書属性の表示管理を行う文書属性表示管理モジュールと、検索結果情報の評価情報を管理し、関連フィードバックが行われた際その文書の表示情報を制御する関連フィードバック管理モジュールとを備えた文書検索装置であって、文書検索での検索結果提示の際に、検索結果の文書間の関連度情報に基づき、関連度の高い文書の情報をまとめて表示させる検索結果情報提示手段（前記各モジュールを含む手段）を備えている。

【0011】(2)：前記(1)の文書検索装置において、前記検索結果情報提示手段は、関連度の高い文書の情報をまとめて表示させる際、多段階に階層的に表示させる機能を備えている。

【0012】(3)：前記(1)の文書検索装置において、前記検索結果情報提示手段は、検索結果の文書間の関連度情報に基づき、上位レベルの文書のみを表示や下位レベルの展開を含むアウトライン操作を行う機能を備えている。

【0013】(4)：前記(1)の文書検索装置において、前記検索結果情報提示手段は、検索結果提示の際に、同一グループに属する複数文書に対し、関連フィードバックの対象とする／しないの指定を一度に行える手段を提供する機能を備えている。

【0014】(5)：前記(1)の文書検索装置において、前記検索結果情報提示手段は、検索結果提示の際に、関連度判定の閾値を、アイテムの密集度に応じて動的に変更する機能を備えている。

【0015】(6)：前記(1)の文書検索装置において、前記検索結果情報提示手段は、関連フィードバックを用いた検索結果提示の際に、関連フィードバックに伴う評

価の向上／下降を表示させる機能を備えている。

【0016】(7)：前記(1)の文書検索装置において、前記検索結果情報提示手段は、関連フィードバックを用いた検索結果提示の際に、関連フィードバックに伴い、評価の向上した文書の表示を優先させる機能を備えている。

【0017】(8)：前記(1)の文書検索装置において、前記検索結果情報提示手段は、検索結果提示の際に、検索結果の文書の文書属性を表示させる機能を備えている。

(9)：前記(7)の文書検索装置において、前記検索結果情報提示手段は、前記検索結果の文書の文書属性を表示させる際、文書中のキーワードといった何らかの観点による特徴素を、背景色の濃淡などの表示特性を用いて表示させる機能を備えている。

【0018】(10)：前記(7)の文書検索装置において、前記検索結果情報提示手段は、前記検索結果の文書の文書属性を表示させる際、文書長及びキーワード密度といった何らかの観点による複数の特徴素を、背景色の濃淡などの複数の表示特性を用いて表示させる機能を備えている。

【0019】(11)：前記(7)の文書検索装置において、前記検索結果情報提示手段は、前記検索結果の文書の文書属性を表示させる際、文書ベクトルの情報要求関連の情報量といった検索キーをも考慮した背景色などの表示特性を用いて表示させる機能を備えている。

【0020】(12)：文書データを格納した文書データベースと、前記文書データベースの文書検索処理を行う検索エンジンと、情報の視覚的操作のためのインターフェース制御を行うグラフィカルユーザインターフェースモジュールと、前記検索エンジンによる検索結果の情報に対して関連情報の統合処理を行う関連統合モジュールと、文書属性の表示管理を行う文書属性表示管理モジュールと、検索結果情報の評価情報を管理し、関連フィードバックが行われた際、その文書の表示情報を制御する関連フィードバック管理モジュールとを備えた文書検索装置に、文書検索での検索結果提示の際に、検索結果の文書間の関連度情報に基づき、関連度の高い文書をまとめて表示させる手順を実行させるためのプログラムを記録したコンピュータ読取可能な記録媒体。

【0021】(作用)前記構成に基づく本発明の作用を説明する。

(a)：前記(1)の作用

グラフィカルユーザインターフェースモジュールは、ユーザより検索キーが与えられると、その検索キーを検索エンジンに渡す。この時、検索エンジンは前記検索キーによるデータベースの検索を行い、検索結果情報を取得する。関連統合モジュールは、前記データベース検索の結果得られた検索結果情報について、予めユーザから指定された関連統合処理を行い、関連統合表示情報を得

る。

【0022】また、文書属性表示管理モジュールは、前記検索結果情報について、文書属性表示のための処理を行い、文書属性表示情報を得る。そして、グラフィカルユーザインターフェースモジュールはディスプレイ装置の画面上に前記各モジュールから受け取った情報を表示してユーザに提示する。

【0023】更に、関連フィードバック管理モジュールは、前記検索結果情報における評価情報を管理し、関連フィードバックが行われた際、表示制御のための情報をグラフィカルユーザインターフェースモジュールへ渡す。この情報を受け取ったグラフィカルユーザインターフェースモジュールは、その情報をディスプレイ装置の画面上に表示してユーザに提示する。

【0024】このような文書検索処理において、検索結果情報提示手段は、文書検索での検索結果提示の際に検索結果の文書間の関連度情報に基づき、関連度の高い文書の情報をまとめて表示させる。

【0025】このようにすれば、通常の検索結果における順位や検索キーとの関連度などの情報に加え、それぞれの文書間の関連情報を含む検索結果情報を提示することができる。そして、文書間の関連度の計算は、検索結果における検索キーとの関連度が高いもののみに限定することができるため、高速な処理が可能になる。更に、検索時に得られる文書データからの情報を活用することで、効率的な検索結果情報や操作手段の提供が可能になる。

【0026】(b)：前記(2)の作用

前記文書検索時において、検索結果情報提示手段は、関連度の高い文書の情報をまとめて表示させる際多段階に階層的に表示させる。このようにすれば、文書間の関係が見やすくなり、検索結果情報が効率的に活用できる。

【0027】(c)：前記(3)の作用

前記文書検索時において、前記検索結果情報提示手段は、検索結果の文書間の関連度情報に基づき、上位レベルの文書のみを表示や下位レベルの展開を含むアウトライン操作を行う。このようにすれば、例えば、上位レベルの文書のみを表示により、検索キーとの関連度の高い文書の概要を大まかに知ることができる。従って、検索漏れを減らすことができる。

【0028】(d)：前記(4)の作用

前記文書検索時において、前記検索結果情報提示手段は、検索結果提示の際に、同一グループに属する複数文書に対し、関連フィードバックの対象とする／しないの指定を一度に行える手段（例えば、画面上にオン／オフボタンを表示し、マウス等によるクリックで指定する手段）を提供する。このようにすれば、効率的な検索結果情報や操作手段の提供が可能になる。

【0029】(e)：前記(5)の作用

前記文書検索時において、前記検索結果情報提示手段

は、検索結果提示の際に関連度判定の閾値をアイテムの密集度に応じて動的に変更する。この場合、文書データベースによっては、検索の結果検索キーとの関連度の高い文書が多く存在することがあり、検索結果の文書が全て一つのグループになってしまうこともある。そのような場合、前記のように、関連度を判定する閾値を動的に変更することで、常に適切な検索結果情報を提示することができる。

【0030】(f)：前記(6)の作用

10 前記文書検索時において、前記検索結果情報提示手段は、関連フィードバックを用いた検索結果提示の際に関連フィードバックに伴う評価の向上／下降を表示させる。このようにすれば、前回の検索結果との順位や関連度といった評価の違いを明確にすることができ、検索結果情報の効率的な利用が可能になる。

【0031】(g)：前記(7)の作用

20 前記文書検索時において、前記検索結果情報提示手段は、関連フィードバックを用いた検索結果提示の際に関連フィードバックに伴い評価の向上した文書の表示を優先させる。このようにすれば、検索漏れを減らし、関連フィードバックによる検索精度向上の効率化を期待できる。

【0032】(h)：前記(8)の作用

前記文書検索時において、前記検索結果情報提示手段は、検索結果提示の際に検索結果の文書の文書属性を表示させる。このようにすれば、通常の検索結果情報に加え、文書に含まれる情報量を提示することで、より効率的な検索結果の利用が可能になる。

【0033】(i)：前記(9)の作用

30 前記文書検索時に検索結果の文書の文書属性を表示させる場合、前記検索結果情報提示手段は、文書中のキーワードといった何らかの観点による特徴素を、背景色の濃淡などの表示特性を用いて表示させる。このようにすれば、特徴素が見やすくなり、検索結果情報の効率的な利用が可能になる。

【0034】(j)：前記(10)の作用

40 前記文書検索時に検索結果の文書の文書属性を表示させる場合、前記検索結果情報提示手段は、文書長及びキーワード密度といった何らかの観点による複数の特徴素を、背景色の濃淡などの複数の表示特性を用いて表示させる。このようにすれば、特徴素が見やすくなり、検索結果情報の効率的な利用が可能になる。

【0035】(k)：前記(11)の作用

50 前記文書検索時に検索結果の文書の文書属性を表示させる場合、前記検索結果情報提示手段は、前記検索結果の文書の文書属性を表示させる際、文書ベクトルの情報要求関連の情報量といった検索キーをも考慮した背景色などの表示特性を用いて表示させる。このようにすれば、特徴素が見やすくなり、検索結果情報の効率的な利用が可能になる。

【0036】(1)：前記(12)の作用

文書検索装置において、前記記録媒体から読み出したプログラムを実行することにより、文書検索での検索結果提示の際に、検索結果の文書間の関連度情報に基づき、関連度の高い文書の情報をまとめて表示させる。

【0037】このようにすれば、通常の実行結果における順位や検索キーとの関連度などの情報に加え、それぞれの文書間の関連情報を含む検索結果情報の提示を得ることができる。そして、文書間の関連度の計算は、検索結果における検索キーとの関連度が高いもののみに限定することができ、高速な処理が可能になる。更に、検索時に得られる文書データからの情報を活用することで、効率的な検索結果情報や操作手段の提供が可能になる。

【0038】

【発明の実施の形態】以下、発明の実施の形態を図面に基づいて詳細に説明する。

§1：文書検索装置の説明・・・図1参照

図1は装置構成図である。以下、図1に基づいて、文書検索装置の構成を説明する。図1に示したように、文書検索装置は、装置本体1と、前記装置本体1に接続されたディスプレイ装置2、キーボード3、ハードディスク装置(HDD)5、プリンタ6、フレキシブルディスク装置(FDD)7等により構成されている。

【0039】また、装置本体1には、文書検索時の各種制御を行うCPU10と、入出力制御を行う入出力制御部11と、メモリ12等が設けられている。この場合、前記ハードディスク装置5には、データベース(後述する)や文書検索処理に必要なプログラムを格納しておき、文書検索を行う場合には、CPU10の制御によりハードディスク装置5内のプログラムを読み出してメモリ12に格納する。そして、CPU10がメモリ12から必要なプログラムを読み出して実行することで、文書検索処理を実行する。なお、前記文書検索装置は、パーソナルコンピュータ、ワークステーション等の各種コンピュータで実現できる。

【0040】§2：文書検索モジュールの説明・・・図2参照

図2は文書検索モジュールの説明図である。前記文書検索装置には、図2に示したような文書検索モジュールが格納されており、これらのモジュール等により文書検索処理を行う。以下、詳細に説明する。

【0041】図2に示したように、前記文書検索装置には、グラフィカル・ユーザ・インターフェースモジュール(Graphical User Interface：以下「GUIモジュール」と記す)21と、文書検索制御モジュール22と、検索エンジン23と、関連統合モジュール24と、文書属性表示管理モジュール25と、関連フィードバック管理モジュール26を備えている。そして、前記モジュールにより文書データベース27の検索処理を行う。

【0042】前記GUIモジュール21は、コンピュータを視覚的に操作できるように考案されたインターフェースモジュールであり、OSやOS上の基本環境として提供されるものである。例えば、ビットマップディスプレイにウィンドウを開き、ドライブやディレクトリ、ファイルをアイコンで表示し、マウスなどのポインティングデバイスで操作する。

【0043】前記文書検索制御モジュール22は、検索エンジン23、関連統合モジュール24、文書属性表示管理モジュール25、関連フィードバック管理モジュール26に対する制御を行うと共に、GUIモジュール21との間で情報交換を行うことで、文書検索処理全体の制御を行うものである。

【0044】前記検索エンジン23は、文書検索制御モジュール22の指示により文書データベース27の検索処理を行い、検索結果を文書検索制御モジュール22へ転送するものである。前記関連統合モジュール24は、文書検索制御モジュール22の指示により検索結果の情報に対して関連する情報の統合処理を行うものである。

【0045】前記文書属性表示管理モジュール25は、文書検索制御モジュール22の指示により、検索結果の情報に対して文書属性の表示管理を行うものである。前記関連フィードバック管理モジュール26は、文書検索制御モジュール22の指示により検索結果の情報における評価情報を管理し、関連フィードバックが行われたら、指定された表示制御のための情報を文書検索制御モジュール22へ転送するものである。

【0046】前記文書データベース27は、文書データ(単に「文書」と記す)を格納したデータベースであり、前記ハードディスク装置5の記憶媒体(又は記録媒体)に記憶されている。なお、前記GUIモジュール21と、文書検索制御モジュール22と、検索エンジン23と、関連統合モジュール24と、文書属性表示管理モジュール25と、関連フィードバック管理モジュール26は、それぞれ文書検索処理用のプログラム(図1のハードディスク装置5から読み出し、メモリ12に格納してCPU10が実行するプログラム)の実行により実現する。

【0047】前記モジュールの動作概要は次の通りである。GUIモジュール21は、ユーザより検索キーが与えられると、その検索キーを文書検索制御モジュール22へ渡す。この検索キーを受け取った文書検索制御モジュール22は、検索エンジン23に指示して文書データベース27の検索を行い、検索結果の情報を取得する。

【0048】文書検索制御モジュール22は、受け取った検索結果情報を関連統合モジュール24及び文書属性表示管理モジュール25へと渡す。関連統合モジュール24は、渡された検索結果情報について予めユーザから指定された関連統合処理を行い、文書検索制御モジュール22へと関連統合表示情報を与える。

【0049】また、文書属性表示管理モジュール25は、前記渡された検索結果情報について文書属性表示のための処理を行い、文書検索制御モジュール22へと文書属性表示情報を与える。そして、文書検索制御モジュール22は前記得られた情報をGUIモジュール21へ渡し、GUIモジュール21はディスプレイ装置2の画面上に前記受け取った情報を表示してユーザに提示する。

【0050】また、関連フィードバック管理モジュール26は、検索結果情報における評価情報を管理し、関連フィードバックが行われた際、表示制御のための情報を文書検索制御モジュール22へ渡す。この情報を受け取った文書検索制御モジュール22は、その情報をGUIモジュール21へ渡し、GUIモジュール21は受け取った情報をディスプレイ装置2の画面上に表示してユーザに提示する。

【0051】§3：具体例による文書検索処理の説明
以下、図に基づいて、前記文書検索装置における文書検索処理の具体例を説明する。なお、以下の例において、検索結果情報の提示はディスプレイ装置2の画面上へ表示することで行う。

【0052】(1)：例1・・・図3、4参照

図3は例1の処理フローチャート(その1)、図4は例1の処理フローチャート(その2)である。例1は、文書検索での検索結果提示の際に、検索結果の文書間の関連度情報に基づき、関連度の高い文書をまとめて表示する例である。以下、図3、4に基づいて例1の処理を説明する。なお、S1～S12は各処理ステップを示す。

【0053】①：全体の処理・・・図3参照

例1は、文書検索での検索結果提示の際に、検索結果の文書間の関連度情報に基づき、関連度の高い文書をまとめて表示する例であり、先ず、文書検索制御モジュール22の指示により、検索エンジン23が文書データベース27を検索し、検索結果情報を得る。

【0054】そして、得られた検索結果情報は文書検索制御モジュール22へ渡された後、関連統合モジュール24へ渡される。このようにして、関連統合モジュール24は検索結果情報、すなわち、検索結果の各文書の情報として、文書ID、検索キーとの関連度、タイトル、対応するベクトル等の情報を得る(S1)。

【0055】次に、関連統合モジュール24は、各文書間の関連度を計算し、関連度が与えられた閾値(予め設定しておく)以上のものを統合する(S2)。そして、統合された文書同士がまとめて表示されるように、検索結果情報を提示する(S3)。

【0056】なお、前記S2の処理では、各文書間の関連度を計算し、関連度が与えられた閾値以上のものを統合するが、この処理の具体例としてはいくつかの処理がある。ここでは、関連度の高い順に簡易的に統合する処理と、クラスタリングに基づき統合する処理について以

下に説明する。

【0057】②：関連度の高い順に簡易的に統合する処理の説明・・・図4のA図参照

図4のA図に示したように、関連度の高い順に簡易的に統合する処理では、次のように処理を行う。先ず、関連統合モジュール24は、パラメータ*i*を初期値1に設定($i \leftarrow 1$)し(S5)、全ての文書について処理が済んだか否かを判断する(S6)。

【0058】その結果、全ての文書について処理が済んでいれば、処理を終了するが、処理が済んでいなければ、*i*位の文書*D_i*はマーク(処理済みか否かを示すマーク)されているか否かを判断する(S7)。その結果、マークされていれば、*i*を更新($i \leftarrow i + 1$)し(S10)、S6の処理から繰り返して行う。しかし、S7の処理で、マークされていない(未処理であれば)、*i* + 1位以下のマークされていない各文書について、*D_i*との関連度を求め、関連度が予め定めた閾値より高いものを*D_i*と統合する(S8)。

【0059】次に、関連統合モジュール24は、前記*D_i*と統合した文書にマークし(S9)、*i*を更新($i \leftarrow i + 1$)し(S10)、S6の処理から繰り返して行う。このようにして、全ての文書について前記処理を行うことで、文書間の関連度の高い順に簡易的に統合することができる。

【0060】③：クラスタリングに基づき統合する処理の説明・・・図4のB図参照

図4のB図に示したように、クラスタリングに基づき統合する処理では次のように処理を行う。先ず、関連統合モジュール24は、検索結果の文書全体を階層型クラスタリングを用いてクラスタ間の距離が閾値(予め設定しておく)を越えるようクラスタC1、C2、・・・C*m*へと分ける(S11)。次に、各クラスタに属する文書を統合する(S12)。このような処理により、クラスタリングに基づき統合する処理を行う。

【0061】(2)：例2・・・図5、6参照

図5は例2の処理フローチャート(その1)、図6は例2の処理フローチャート(その2)である。例2は、前記関連度の高い文書をまとめて表示する際、多段階に階層的に表示する例である。以下、図5、6に基づいて例2の処理を説明する。なお、S21～S39は各処理ステップを示す。

【0062】①：全体の処理・・・図5のA図参照

例2の処理は、前記例1において、関連度の高い文書をまとめて表示する際、多段階に階層的に表示する例であり、図5のA図に示したように、先ず、関連統合モジュール24は、検索エンジン23より、文書検索制御モジュール22を介して検索結果の各文書の情報(文書ID、検索キーとの関連度、タイトル、対応するベクトル)を得る(S21)。

【0063】次に、関連統合モジュール24は、各文書

間の関連度を計算し、関連度が与えられた閾値以上のものを階層的に統合する(S22)。そして、前記統合した情報を文書検索制御モジュール22を介してGUIモジュール21へ渡し、該GUIモジュール21により画面に表示することで、統合された文書同士が纏まって表示されるように、検索結果を提示する(S23)。

【0064】このようにして、関連度の高い文書をまとめて表示する際、多段階に階層的に表示することができる。なお、前記関連度の高いものを階層的に統合する処理には幾つかのやり方がある。ここでは、クラスタリングに基づき統合する処理と、関連度の高い順に簡易的に統合する処理とについて以下に説明する。

【0065】②：クラスタリングに基づき統合する処理の説明・・・図5のB図参照

図5のB図に示したように、クラスタリングに基づき統合する処理は次のようにして処理を行う。まず、関連統合モジュール24は、検索結果の文書全体を、階層型クラスタリングを用いてクラスタ間の距離が予め定めた閾値を越えるようクラスタC1、C2、・・・Cmへと分ける。その際、クラスタリング過程の階層情報をメモリ12に保存する(S25)。そして、関連統合モジュール24は、各クラスタの階層情報に基づいて、文書を統合する(S26)。このようにして、クラスタリングに基づき統合する処理を行う。

【0066】③：関連度の高い順に統合する処理の説明・・・図6のC図参照

図6のC図に示したように、前記関連度の高い順に統合する処理は次のようにして行う。まず、関連統合モジュール24は、パラメータ*i*を初期化($i \leftarrow 1$)し(S31)、全ての文書について処理が済んだか否かを判断する(S32)。その結果、全ての処理が済んでいれば、処理を終了するが、全ての処理が済んでいなければ、*i*位の文書*D_i*はマーク(処理済みか否かを示すマーク)されているか否かを判断する(S33)。

【0067】その結果、*i*位の文書*D_i*がマークされていれば、*i*を更新($i \leftarrow 1$)し(S35)、S32の処理から繰り返して行う。しかし、*i*位の文書*D_i*がマークされていなければ、統合処理integ(*D_i*) (詳細は後述する)を行い(S34)、*i*を更新($i \leftarrow 1$)し(S35)、S32の処理から繰り返して行う。

【0068】④：統合処理integ(*D_i*)・・・図6のD図参照

図6のD図に示したように、前記S34の処理で行う統合処理integ(*D_i*)は次のようにして行う。まず、関連統合モジュール24は、*i*+1位以下のマークされていない各文書について、*D_i*との関連度を求め、関連度が予め定めた閾値より高いものを*D_i*と統合する(S37)。

【0069】次に、*D_i*と統合した文書にマーク(処理済みか否かを示すマーク)し(S38)、*D_i*(の直

下)に統合された各文書*D_j*について、統合処理integ(*D_j*)を再帰的に実行する(S39)。このようにして、前記統合処理integ(*D_i*)を行う。

【0070】(3)：例3・・・図7参照

図7は例3の処理フローチャートである。例3は、文書検索時において、検索結果の文書間の関連度情報に基づき、上位レベルの文書のみの表示や下位レベルの展開を含むアウトライン操作を提供する例である。以下、図7に基づいて例3の処理を説明する。なお、S41～S46は各処理ステップを示す。

【0071】①：全体の処理・・・図7のA図参照

例3は、例1の文書検索処理において、検索結果の文書間の関連度情報に基づき、上位レベルの文書のみの表示や、下位レベルの展開といったアウトライン操作を提供する例であり、図7のA図に示したように、まず、関連統合モジュール24は、検索エンジン23より、文書検索制御モジュール22を介して検索結果の各文書の情報を得る(S41)。この場合の検索結果の各文書の情報には、文書ID、検索キーとの関連度、タイトル、対応するベクトル等を含む。

【0072】次に、関連統合モジュール24は、各文書間の関連度を計算し、関連度に基づき文書間の階層性を抽出する様に統合する(S42)。そして、抽出された階層性に基づき、ユーザ指示に基づき、アウトライン操作を実行する(S43)。このようにして、検索結果の文書間の関連度情報に基づき、上位レベルの文書のみの表示や、下位レベルの展開といったアウトライン操作を提供する。

【0073】なお、前記アウトライン操作には様々なものがあるが、基本的には各文書に対応するノード(文書)の展開と、ノードの統合の組み合わせで実現される。この場合の処理は次のようになっている。

【0074】②：ノードの展開の説明・・・図7のB図参照

図7のB図に示したように、ノードの展開では、或るノード(文書)の直下のノードを表示に追加する(S45)。このようにしてノードの展開を行う。

【0075】③：ノードの統合の説明・・・図7のC図参照

図7のC図に示したように、ノードの統合は、あるノード(文書)の下位のノードの表示を表示から削除する(S46)。このようにしてノードの統合を行う。

【0076】(4)：例4・・・図無し

例4は、例1の文書検索処理において、検索結果の提示の際に、同一グループのオン/オフ指定を一度に行えるようにした例である。例4では、関連フィードバック指定の際のGUIとして、検索結果の文書間の関連度情報に基づく階層性を利用して、同じグループに属する文書を一度にオン/オフ(関連フィードバックの対象とする/しない)指定を行うボタンを設けることで、関連フィ

10

20

30

40

50

ードバック指定を効率化する。

【0077】なお、前記ボタンは、例えば、表示画面上に表示されたクリックボタン（マウスでクリック可能なボタン）であり、このボタンを使って、複数文書を一度に指定できるようにするものである。

【0078】(5)：例5・・・図8参照

図8は例5の処理フローチャートである。例5は、前記文書検索処理において、検索結果提示の際に、関連度判定の閾値を、アイテムの密集度に応じて動的に変更する例である。以下、図8に基づいて例5の処理を説明する。なお、S51～S55は各処理ステップを示す。

【0079】例5は、例1の文書検索処理において、検索結果提示の際に、関連度判定の閾値を、アイテムの密集度に応じて動的に変更する例であり、図8に示したように、関連統合モジュール24は、先ず、検索エンジン23より文書検索制御モジュール22を介して検索結果の各文書の情報を得る（S51）。この場合の検索結果の各文書の情報には、文書ID、検索キーとの関連度、タイトル、対応するベクトル等を含む。

【0080】次に、関連統合モジュール24は、各文書間の関連度を計算し、関連度に基づき文書間の階層性を抽出する様に統合する（S52）。この場合、関連度の計算結果は、再試行の時のためにメモリ12に保存しておくという。次に、関連統合モジュール24は、統合結果のグループ数が予め決めたグループ数より少ないか否かを判断する（S53）。

【0081】その結果、統合結果のグループ数が、予め決めたグループ数より少なければ、関連度判定の閾値を上げ（S54）、前記閾値の値が、予め決めた規定値以下か否かを判断する（S55）。その結果、予め決めた規定値以下ならば、S52の処理から繰り返して行い、予め決めた規定値以下でなければ、処理を終了する。また、前記S53の処理で、統合結果のグループ数が、予め決めたグループ数より少なくなければ、処理を終了する。

【0082】(6)：例6・・・図9のB図参照

図9のB図は例6の処理フローチャートである。例6は、例1の文書検索処理において、関連フィードバックを用いた検索結果提示の際に、関連フィードバックに伴う評価の向上／下降を表示する例である。以下、図9のB図に基づいて例6の処理を説明する。なお、S65～S67は各処理ステップを示す。

【0083】例6は、例1の文書検索処理において、関連フィードバックを用いた検索結果提示の際に、関連フィードバックに伴う評価の向上／下降を表示する例であり、図9のB図に示したように、関連フィードバック管理モジュール26は、検索結果における各文書について、以下の処理を行う（S65）。

【0084】先ず、文書IDを元に、前回の検索結果との評価（その文書の順位や関連度）の比較を行う（S6

6）。次に、評価の向上／下降情報を表示するための処理を行う（S67）。

【0085】この場合、関連フィードバック管理モジュール26は、前記評価の向上／下降情報を作成し、文書検索制御モジュール22へ渡す。この情報を受け取った文書検索制御モジュール22はその情報をGUIモジュール21へ送り、このGUIモジュール21がディスプレイ装置2の画面上に前記情報を表示させる。

【0086】(7)：例7・・・図10参照

図10は例7の処理フローチャートである。例7は、例1の文書検索処理において、関連フィードバックを用いた検索結果提示の際に関連フィードバックに伴い、評価の向上した文書の表示を優先させる例である。以下、図10に基づいて例7の処理を説明する。なお、S71～S74は各処理ステップを示す。

【0087】図10に示したように、関連フィードバック管理モジュール26は、検索結果における各文書について、以下の処理を行う（S71）。先ず、文書IDを元に、前回の検索結果との評価（その文書の順位や関連度）の比較を行い（S72）、評価の下降が、予め決めた値より大きいのか否かを判断する（S73）。

【0088】その結果、評価の下降が、予め決めた値より大きければ、その文書の表示を抑制し（S74）し、処理を終了する。また、評価の下降が予め決めた値より小さくなければ、そのまま処理を終了する。このようにして、関連フィードバックを用いた検索結果提示の際に、関連フィードバックに伴い、評価の向上した文書の表示を優先させる処理を行う。

【0089】(8)：例8～11・・・図9のA図参照

図9のA図は例8～例11の処理フローチャートである。以下、図9のA図に基づいて例8～例11の処理を説明する。なお、S61、S62は各処理ステップを示す。例8は、例1の文書検索処理において、検索結果提示の際に検索結果の文書の文書属性を提示する例である。例9は、例8において、文書中のキーワードといった何らかの観点による特徴素を、背景色の濃淡などの表示特性を用いて提示する例である。

【0090】例10は、例8において、文書長及びキーワード密度といった何らかの観点による複数の特徴素を背景色の濃淡などの複数の表示特性を用いて提示する例である。例11は、例8において、文書ベクトルの情報要求関連の情報量といった検索キーをも考慮した背景色などの表示特性を用いた提示する例である。

【0091】図9のA図に示したように、文書属性表示管理モジュール25は、文書検索制御モジュール22を介して検索エンジン23により検索結果の各文書の情報（文書ID、検索キーとの関連度、タイトル、対応するベクトル、属性情報）を得る（S61）。そして、検索結果の各文書について（必要ならば適宜計算を行い）文書属性を提示する。

§4: 検索結果情報例表示例の説明・・・図11～図15参照

図11～図15は、検索結果情報表示例(その1)～(その5)である。以下、図11～図15に基づいて前記各例の処理による検索結果情報表示例を説明する。なお、以下に説明する各例は、それぞれ前記例1～11に対応している。

【0092】(1): 例1・・・図11のA図参照

図11のA図は、前記例1の検索処理により得られた結果の表示例(関連統合表示)である。この例では、前記従来例と同様に、数百件の文書(論文)からの検索結果であり、検索キーとして「概念学習」を用いたものである。

【0093】前記従来例に対し、例1では、関連統合モジュール24で文書間の関連度を求め、関連度が高いもの同士をまとめて表示すると、図11のA図のように、従来例の検索結果における順位や検索キーとの関連度などの情報に加え、それぞれの文書間の関連情報を含む検索結果提示ができる。図11の検索結果情報では、各文書毎に、順位、関連度、文書ID、文書タイトルの順に表示されている。

【0094】この例では、順位=1は、関連度=0.84、文書ID=92-89-03であり、文書タイトルは、「名詞述語文からの概念学習システム-比喻的理解と学習-」となっている。順位=2は、関連度=0.78、文書ID=90-79-04であり、文書タイトルは、「名詞述語文における意味概念の学習」となっている。

【0095】順位=3は、関連度=0.74、文書ID=91-83-19であり、文書タイトルは、「自然言語概念学習に関する考察」となっている。順位=4は、関連度=0.72、文書ID=93-93-06であり、文書タイトルは、「名詞述語文からの概念学習システム-継承と経路情報の処理-」となっている。

【0096】順位=5は、関連度=0.65、文書ID=92-92-11であり、文書タイトルは、「ハイパー・シソーラスとその学習」となっている。順位=6は、関連度=0.56、文書ID=90-78-06であり、文書タイトルは、「記述量圧縮の観点から見た概念体系の構築」となっている。

【0097】順位=7は、関連度=0.54、文書ID=91-85-09であり、文書タイトルは、「概念の動的管理システム」となっている。順位=8は、関連度=0.50、文書ID=91-83-14であり、文書タイトルは、「語彙概念と知識について」となっている。

【0098】順位=9は、関連度=0.45、文書ID=90-75-07であり、文書タイトルは、「自然言語意味学習の試み」となっている。順位=10は、関連度=0.45、文書ID=90-75-12であり、文

書タイトルは、「概念辞書の作成」となっている。

【0099】この場合、文書間の関連度の計算は、検索結果における検索キーとの関連度が高いもののみに限定することができるため、高速な処理が可能になる。なお、この例1では、次の順序で処理を行い、図示の情報を得た。

【0100】①: 順位1との関連度の高い(或る閾値以上の)ものを求める。その結果、順位2、4の検索結果情報(順位1に対して、順位2、4は下位レベルの情報)を得る。

【0101】②: 順位2はマーク済み(処理済み)なので、次に、順位3との関連度の高い(或る閾値以上の)ものを求める。その結果、順位9の検索結果情報(順位3に対して、順位9は下位レベルの情報)を得る。

【0102】③: 順位4はマーク済み(処理済み)なので、次に、順位5との関連度の高い(或る閾値以上の)ものを求める。その結果、関連度の高いものは得られなかった。

【0103】④: 順位6との関連度の高い(或る閾値以上の)ものを求める。その結果、関連度の高いものは得られなかった。

⑤: 順位7との関連度の高い(或る閾値以上の)ものを求める。その結果、順位10の検索結果情報(順位7に対して、順位10は下位レベルの情報)を得る。

【0104】⑥: 順位8との関連度の高い(或る閾値以上の)ものを求める。その結果、関連度の高いものは得られなかった。

(2): 例2・・・図11のB図参照

図11のB図は、前記例2の検索処理により得られた結果の表示例(関連統合表示)である。この例2では、関連度の高い文書をまとめて表示する際、多段階に階層的に表示する。すなわち、1レベルの階層の代わりに再帰的な処理を行うと、図示のような多段階の階層的表示(階層的な関連統合表示)を得る。この場合、図示した各順位の検索結果情報は、前記例1と同じである。なお、この例2では、次の順序で処理を行い、図示の情報を得た。

【0105】①: 順位1との関連度の高い(或る閾値以上の)ものを求める。その結果、順位2、4の検索結果情報を得る。

②: 順位2との関連度の高い(或る閾値以上の)ものを求める。その結果、順位7の検索結果情報を得る。

【0106】③: 順位7との関連度の高い(或る閾値以上の)ものを求める。その結果、順位10の検索結果情報を得る。

④: 順位4との関連度の高い(或る閾値以上の)ものを求める。その結果、関連度の高いものは得られなかった。

【0107】⑤: 順位3との関連度の高い(或る閾値以上の)ものを求める。その結果、順位9の検索結果情報を得る。以下、同様にして、図示の検索結果情報を得た。

(3): 例3・・・図12参照

図12のA図、B図は、前記例3の検索処理により得られた結果の表示例(アウトライン処理による表示例)であり、A図は上位レベルの文書のみを表示例、B図は下位レベル文書の展開表示例である。

【0108】例3では、検索結果における文書間の関連性による階層構造に注目することで、上位レベルの文書のみを表示(図12のA図参照)や、下位レベル文書の展開(図12のB図参照)といった効率の良いアウトライン操作が可能になる。

【0109】そして、例えば、上位レベルの文書のみを表示により、検索キーと関連度の高い文書の概要を、より大まかに知ることができ、検索漏れの可能性を減らすことができる。図12の検索結果情報では、各文書毎に、順位、関連度、文書ID、文書タイトルの順に表示されている。

【0110】図12のA図の例では、図11のA図に示した表示例において、上位レベルの文書のみを表示している。すなわち、図11のA図において、順位1に対して、順位2、4は下位レベルの情報であり、順位3に対して、順位9は下位レベルの情報であり、順位7に対して、順位10は下位レベルの情報である。そこで、前記下位レベルである、順位2、4、9、10の情報を無くし、上位レベルの情報のみを表示すると、図12のA図に示した例3の情報を得る。

【0111】従って、例3の図12のA図に示した例では、順位1、3、5、6、7、8の各情報が表示されると共に、新たに、順位11、12、14、15の各情報が表示されている。なお、この場合、下位レベル文書の数(自分を含む)も表示することで、下位レベル文書の

情報も提示している。

【0112】すなわち、各文書について、順位1では下位レベル文書の数=3、順位3では下位レベル文書の数=2、順位5では下位レベル文書の数=1、順位6では下位レベル文書の数=1、順位7では下位レベル文書の数=2、順位8では下位レベル文書の数=1、順位11では下位レベル文書の数=1、順位12では下位レベル文書の数=2、順位14では下位レベル文書の数=1、順位15では下位レベル文書の数=1となっている。

【0113】また、図12のB図に示した例3は、図11のA図に示した表示例において、上位レベルの文書のみを表示しているだけでなく、順位7の文書についてのみ、下位レベル文書の展開(順位7の下位レベルが順位10として表示されている)が行われている。

【0114】(4): 例5・・・図13参照

図13は検索結果情報表示例(その3)であり、前記例5の検索処理により得られた結果の表示例(動的な関連度閾値の変更による表示例)である。図13のA図は閾値を或る値に設定した場合の表示例、B図は前記閾値を上げた場合の表示例を表す。

【0115】ところで、文書データベースによっては、検索の結果検索キーと関連度の高い文書が多く存在することがあり、検索結果の文書が全て一つのグループになってしまうことがある。そのような場合、関連度を判定する閾値を動的に変更することで、常に適切な検索結果情報の提示を得ることができる。このように動的に関連度の閾値を変更して検索結果を得た例が図13に示してある。

【0116】この例では、検索キーを「機械翻訳」として、前記文書データベース27を検索した結果の情報例であり、順位、関連度、文書ID、文書タイトルの順に表示されている。図13のA図に示した例では、順位1の文書は、順位=1、関連度=0.91、文書ID=91-82-05、文書タイトル=用例主導型機械翻訳となっており、この順位1の文書の下位レベル文書として、順位2~10の各文書が並んでいる。

【0117】すなわち、順位2~10は全て順位1に関連した文書として表示されているが、これらの下位レベルの表示は、密集度が高過ぎる。そこで、関連度数の閾値を上げて検索を行うと、図13のB図のような検索結果情報の表示が得られる。

【0118】この例では、閾値を上げたことで、順位1の下位レベルは順位2のみとなる。また、順位3の下位レベルは順位5のみとなり、順位4の下位レベルは順位9のみとなり、順位6の下位レベルは順位7、8、10となる。このように、検索結果の提示の際に、関連度判定の閾値をアイテムの密集度に応じて動的に変更することで、常に、適切な検索結果情報の提示が可能になる。

【0119】(5): 例6~例11の説明・・・図14、図15参照

図14は検索結果情報表示例(その4)、図15は検索結果情報表示例(その5)である。検索結果提示においては、検索キーとの関連度の強さのみでなく、検索された文書に含まれる情報量もユーザにとって重要である。すなわち、関連度は、「文書Aの方が少し高いが、文書Bの方が内容が多そうなので、先ず、文書Bから目を通そう」ということがあるからである。

【0120】そこで、通常の検索結果に加え、文書に含まれる情報量を提示することで、より効率的な検索結果の利用が可能になる(例8)。その際の提示に用いる文書情報量や、提示方法としては、①: 文書中のキーワードの量(キーワードの総数など)を、背景色の濃淡で提示する(例9)。②: 文書長(文書単語数、或いは文書文字数)を背景色の長さで、キーワード密度(キーワード量/文書長)を背景色の濃淡で表示する(例10)。③: 文書ベクトルの情報要求関連の情報量(キーベクトル方向成分)を提示する(例11)などが考えられる。

【0121】実際の検索では、或る検索キーで検索を行った後、関連フィードバックを繰り返し行うことで、よ

り適切に検索結果を得るための絞り込みを行うことが多い。関連フィードバックを行うと、通常の場合、前回とは異なる検索結果を得る。その際、どこがどう変化したのか分かりやすく提示することができれば、検索結果の有効利用につながる。

【0122】本発明では、関連フィードバック情報（或いは、評価情報）を用いることで、前回の検索結果との順位や関連度といった評価の違いを求め、それを検索結果と一緒に提示する（例6）。関連フィードバックにより評価が下がる文書は、例えば、順位が上の方でも、その検索においては、あまり重要ではないと考えられる。

【0123】逆に、順位はそれほど高くなくても、関連フィードバックにより評価が上がる文書は重要である可能性がある。そこで、評価の向上した文書の表示を優先させることで検索漏れを減らし、関連フィードバックによる検索精度向上の効率化を期待することができる（例7）。前記関連フィードバックに伴う評価の変化の提示例を図14、図15に基づいて説明する。

【0124】図14に示した検索結果情報表示例は、検索キーを「曖昧」として文書データベース27を検索した結果の情報を表示した例である。この表示例でも前記表示例と同様に、各文書について、順位、関連度、文書ID、文書タイトルの順に表示されている。このような検索結果表示情報に対し、図示の*印の文書（順位が1、2、3、5、7、8の各文書）を対象として関連フィードバックを行うと、図15のA図に示した例6の検索結果情報表示例が得られた。

【0125】この例6の表示例では、前回の検索結果との順位や、関連度といった評価の違いを求め、それを検索結果と一緒に提示する。この場合、任意の整数nに対し、「↑n」はn位評価向上を表し、「↓n」はn位評価下降を表し、「-0」は評価変化無しを表す。

【0126】また、図14の検索結果表示情報に対し、図示の*印の文書（順位が1、2、3、5、7、8の各文書）を対象として関連フィードバックを行い、評価向上文書の優先表示を行うと、図15のB図に示した例7の検索結果情報表示例が得られた。

【0127】この場合、「↓3」以上の表示を抑制（「↓3」より悪いものを表示しない）しているので、図15のA図に示した順位8、9の文書は表示が抑制され（表示されない）、新たに、順位10、11、12の文書が表示されている。このようにすれば、検索漏れを減らし、関連フィードバックによる検索精度向上の効率化を期待することができる。

【0128】§5：記録媒体とプログラムの説明
前記文書検索装置が行う処理、すなわち、文書データベース27の検索処理を行う検索エンジン23と、文書検索時の制御を行う文書検索制御モジュール22と、情報の視覚的操作のためのインターフェース制御を行うGUIモジュール21と、検索結果情報の関連統合処理を行

う関連統合モジュール24と、文書属性の表示管理を行う文書属性表示管理モジュール25と、検索結果情報の評価情報を管理し、関連フィードバックが行われた際、その文書の表示情報を制御する関連フィードバック管理モジュール26により行う文書検索処理は、文書検索装置内のCPU10がメモリ12に格納したプログラムを実行することにより、次のようにして実現する。

【0129】前記文書検索装置に設けたハードディスク装置5の記録媒体（ハードディスク）に、文書データベース27や、前記文書検索処理を実現するためのプログラム、各種パラメータ等のデータを格納しておく。そして、文書検索処理を行う場合は、CPU10の制御によりハードディスク装置5の記録媒体に格納されている前記プログラム等を読み出して装置本体1内のメモリ12に取り込む。

【0130】その後、CPU10がメモリ12に格納してあるプログラムの内、必要なプログラムから順次読み出して実行することにより、前記文書検索装置の文書検索処理を行う。なお、前記ハードディスク装置5の記録媒体に格納するプログラムは、次のようにして記録（記憶）する。

【0131】①：フレキシブルディスク（フロッピーディスク）に格納されているプログラム（他の装置で作成したプログラムデータ）を、フレキシブルディスク装置（FDD）7により読み取り、ハードディスク装置5の記録媒体（ハードディスク）に格納する。

【0132】②：光磁気ディスク、或いはCD-ROM等の記憶媒体に格納されているデータを、前記装置本体1に設けたドライブ装置（図示省略）により読み取り、ハードディスク装置5の記録媒体（ハードディスク）に格納する。

【0133】③：LAN等の通信回線を介して他の装置から伝送されたデータを前記装置本体1で受信し、そのデータをハードディスク装置5の記録媒体（ハードディスク）に格納する。

【0134】

【発明の効果】以上説明したように、本発明によれば次のような効果がある。

(1)：文書検索時に得られる文書データからの情報を活用することで、効率的な検索結果提示や、操作手段の提供が可能になる。

【0135】(2)：大規模な商用オンラインデータベースや、全文検索システムにおいても、検索精度や検索システムの使い勝手が良くなり、効率良く文書検索を行うことが可能になる。

【0136】前記効果の他、各請求項に対応して次のような効果がある。

(3)：請求項1では、グラフィカルユーザインターフェースモジュールはユーザより検索キーが与えられるとその検索キーを検索エンジンに渡し、検索エンジンは検索

10

20

30

40

50

キーによるデータベースの検索を行い検索結果情報取得する。関連統合モジュールはデータベース検索の結果得られた検索結果情報について、予めユーザから指定された関連統合処理を行い関連統合表示情報を得る。

【0137】また、文書属性表示管理モジュールは検索結果情報について文書属性表示のための処理を行い文書属性表示情報を得る。そして、グラフィカルユーザインターフェースモジュールはディスプレイ装置の画面上に各モジュールから受け取った情報を表示してユーザに提示する。

【0138】更に、関連フィードバック管理モジュールは検索結果情報における評価情報を管理し、関連フィードバックが行われた表示制御のための情報をグラフィカルユーザインターフェースモジュールへ渡し、その情報をディスプレイ装置の画面上に表示してユーザに提示する。このような文書検索処理において、検索結果情報提示手段は文書検索での検索結果提示の際に検索結果の文書間の関連度情報に基づき、関連度の高い文書の情報をまとめて表示させる。

【0139】このようにすれば、通常の検索結果における順位や検索キーとの関連度などの情報に加え、それぞれの文書間の関連情報を含む検索結果情報を提示することができる。そして、文書間の関連度の計算は、検索結果における検索キーとの関連度が高いもののみに限定することができるため、高速な処理が可能になる。更に、検索時に得られる文書データからの情報を活用することで、効率的な検索結果情報や操作手段の提供が可能になる。

【0140】(4)：請求項2では、文書検索時において、検索結果情報提示手段は関連度の高い文書の情報をまとめて表示させる際多段階に階層的に表示させる。このようにすれば文書間の関係が見やすくなり、検索結果情報が効率的に活用できる。

【0141】(5)：請求項3では、文書検索時において、検索結果情報提示手段は検索結果の文書間の関連度情報に基づき上位レベルの文書のみ表示や下位レベルの展開を含むアウトライン操作を行う。このようにすれば、例えば、上位レベルの文書のみ表示により、検索キーとの関連度の高い文書の概要を大まかに知ることができる。従って、検索漏れを減らすことができる。

【0142】(6)：請求項4では、文書検索時において、前記検索結果情報提示手段は検索結果提示の際に、同一グループに属する複数文書に対し関連フィードバックの対象とする／しないの指定を一度に行える手段を提供する。このようにすれば、効率的な検索結果情報や操作手段の提供が可能になる。

【0143】(7)：請求項5では、文書検索時において、前記検索結果情報提示手段は検索結果提示の際に関連度判定の閾値をアイテムの密集度に応じて動的に変更する。この場合、文書データベースによっては検索の結

果検索キーとの関連度の高い文書が多く存在することがあり、検索結果の文書が全て一つのグループになってしまうこともある。そのような場合、関連度を判定する閾値を動的に変更することで、常に適切な検索結果情報を提示することができる。

【0144】(8)：請求項6では、文書検索装置において、記録媒体から読み出したプログラムを実行することにより、文書検索での検索結果提示の際に、検索結果の文書間の関連度情報に基づき、関連度の高い文書の情報をまとめて表示させる。

【0145】このようにすれば、通常の検索結果における順位や検索キーとの関連度などの情報に加え、それぞれの文書間の関連情報を含む検索結果情報の提示を得ることができる。そして、文書間の関連度の計算は、検索結果における検索キーとの関連度が高いもののみに限定することができるため、高速な処理が可能になる。更に、検索時に得られる文書データからの情報を活用することで、効率的な検索結果情報や操作手段の提供が可能になる。

20 【図面の簡単な説明】

【図1】本発明の実施の形態における装置構成図である。

【図2】本発明の実施の形態における文書検索モジュールの説明図である。

【図3】本発明の実施の形態における例1の処理フローチャート（その1）である。

【図4】本発明の実施の形態における例1の処理フローチャート（その2）である。

30 【図5】本発明の実施の形態における例2の処理フローチャート（その1）である。

【図6】本発明の実施の形態における例2の処理フローチャート（その2）である。

【図7】本発明の実施の形態における例3の処理フローチャートである。

【図8】本発明の実施の形態における例5の処理フローチャートである。

【図9】本発明の実施の形態における例6、例8～11の処理フローチャートである。

40 【図10】本発明の実施の形態における例7の処理フローチャートである。

【図11】本発明の実施の形態における検索結果情報表示例（その1）である。

【図12】本発明の実施の形態における検索結果情報表示例（その2）である。

【図13】本発明の実施の形態における検索結果情報表示例（その3）である。

【図14】本発明の実施の形態における検索結果情報表示例（その4）である。

50 【図15】本発明の実施の形態における検索結果情報表示例（その5）である。

【図16】従来の検索結果情報表示例である。

【符号の説明】

- 1 装置本体
- 2 ディスプレイ装置
- 3 キーボード
- 4 マウス
- 5 ハードディスク装置 (HDD)
- 6 プリンタ
- 7 フレキシブルディスク装置 (フロッピーディスクドライブ: FDD)
- 10 CPU (中央処理装置)

* 11 入出力制御部

12 メモリ

21 GUIモジュール (グラフィカルユーザインターフェース)

22 文書検索制御モジュール

23 検索エンジン

24 関連統合モジュール

25 文書属性表示管理モジュール

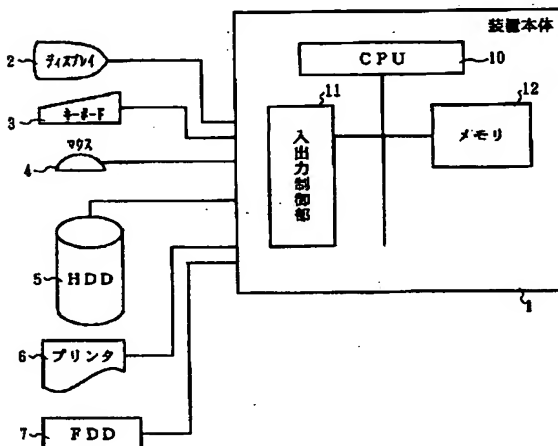
26 関連フィードバック管理モジュール

27 文書データベース

*

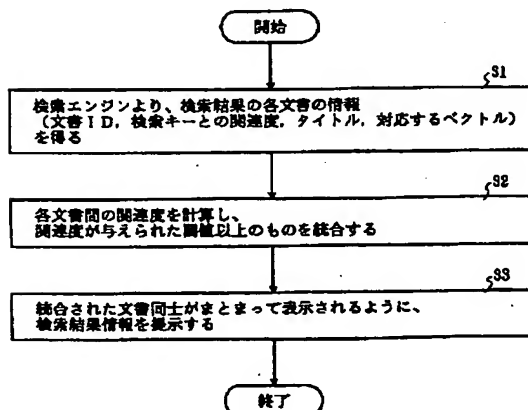
【図1】

装置構成図



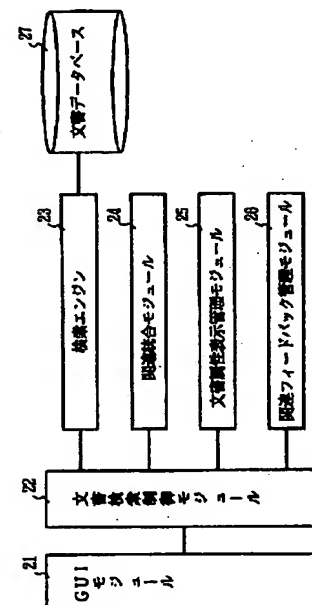
【図3】

例1の処理フローチャート (その1)



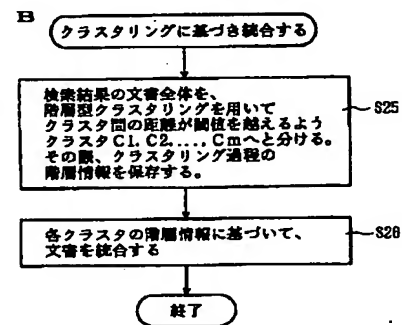
【図2】

文書検索モジュールの説明図



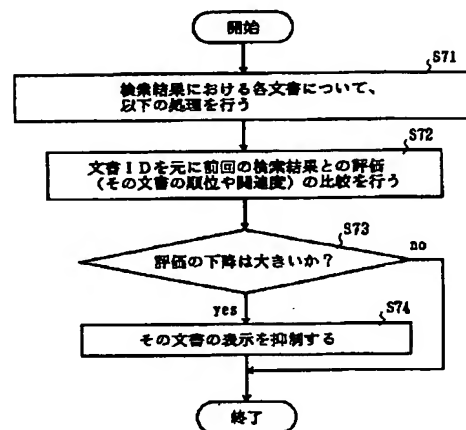
【圖5】

例 2 の処理フローチャート (その 1)



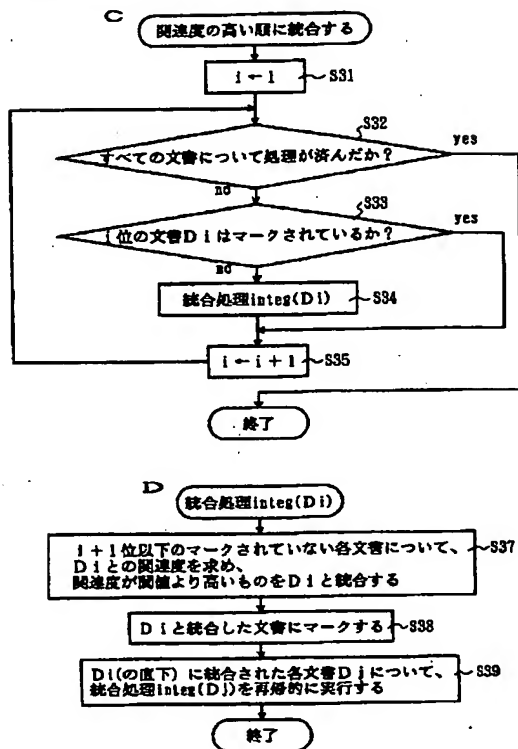
【圖8】

例 7 の処理フローチャート



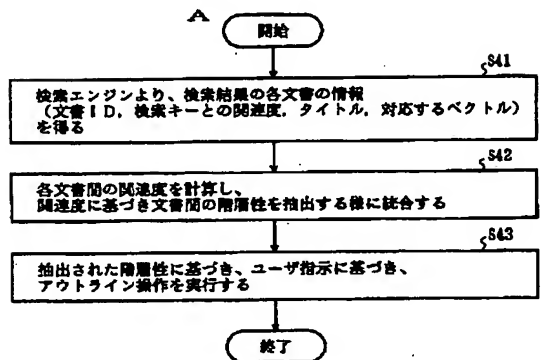
【図6】

例2の処理フローチャート（その2）

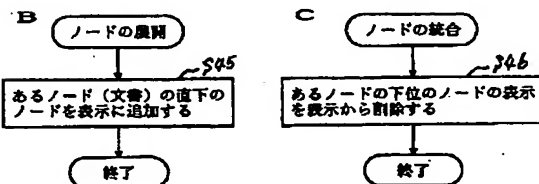


【図7】

例3の処理フローチャート

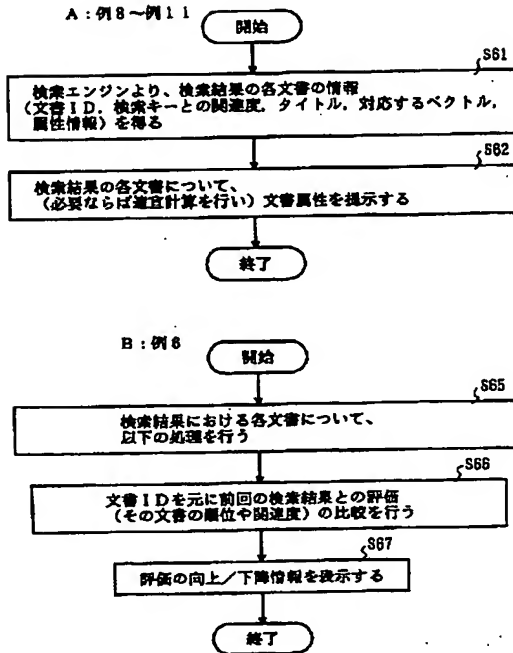


アウトライン操作にはさまざまなものがあるが、基本的には各文書に対応するノードの展開と統合の組合せで実現される。



【図9】

例6、例8～11の処理フローチャート



【図11】

検索結果情報表示例(その1)

A: 例1 検索キー「関心学」 (属性)(関連度)(文書ID) (文書タイトル) (文書属性)

1 (0.84): 92-89-03 各関心学からの関心学システム-出版の関心と学習 -
2 (0.78): 90-79-04 各関心学からの関心学システム-出版の関心と学習 -
3 (0.73): 93-82-06 各関心学からの関心学システム-出版の関心と学習 -
4 (0.74): 91-83-18 各関心学からの関心学システム-出版の関心と学習 -
5 (0.65): 92-82-11 各関心学からの関心学システム-出版の関心と学習 -
6 (0.58): 90-79-06 各関心学からの関心学システム-出版の関心と学習 -
7 (0.54): 91-83-09 各関心学からの関心学システム-出版の関心と学習 -
8 (0.50): 90-79-13 各関心学からの関心学システム-出版の関心と学習 -
9 (0.50): 91-83-14 各関心学からの関心学システム-出版の関心と学習 -

B: 例2 検索キー「関心学」 (属性)(関連度)(文書ID) (文書タイトル) (文書属性)

1 (0.84): 92-89-03 各関心学からの関心学システム-出版の関心と学習 -
2 (0.78): 90-79-04 各関心学からの関心学システム-出版の関心と学習 -
3 (0.73): 93-82-06 各関心学からの関心学システム-出版の関心と学習 -
4 (0.74): 91-83-18 各関心学からの関心学システム-出版の関心と学習 -
5 (0.65): 92-82-11 各関心学からの関心学システム-出版の関心と学習 -
6 (0.58): 90-79-06 各関心学からの関心学システム-出版の関心と学習 -
7 (0.54): 91-83-09 各関心学からの関心学システム-出版の関心と学習 -
8 (0.50): 90-79-13 各関心学からの関心学システム-出版の関心と学習 -
9 (0.50): 91-83-14 各関心学からの関心学システム-出版の関心と学習 -

【図14】

検索結果情報表示例(その4)

検索キー「意味」

↓

1(0.95):90-75-01 Discrimination network上での音韻的曖昧性問題について
 2(0.93):92-81-08 曖昧性を用いた曖昧性問題
 3(0.77):90-80-04 文脈を用いた曖昧性問題の「アプローチ」
 4(0.74):81-82-11 アスペクトをもとにした文脈解析
 5(0.73):81-84-02 曖昧性問題のための曖昧性問題の導入した日本語解析
 6(0.69):92-82-10 曖昧性問題のための曖昧性問題の導入した日本語解析
 7(0.72):92-87-01 曖昧性問題のための曖昧性問題の導入した日本語解析
 8(0.67):90-75-01 曖昧性問題のための曖昧性問題の導入した日本語解析
 9(0.67):90-75-01 曖昧性問題のための曖昧性問題の導入した日本語解析
 10(0.63):92-87-02 曖昧性問題のための曖昧性問題の導入した日本語解析

↑

(順位)(曖昧性)(文書ID) (文書タイトル)

【図15】

検索結果情報表示例(その5)

A:例6 ↓ 「*」についての関連ワードバッグ

↓

04 1(0.95):90-75-01 Discrimination network上での音韻的曖昧性問題について
 03 2(0.93):92-81-08 曖昧性を用いた曖昧性問題
 02 3(0.77):90-80-04 文脈を用いた曖昧性問題の「アプローチ」
 01 4(0.74):81-82-11 アスペクトをもとにした文脈解析
 00 5(0.73):81-84-02 曖昧性問題のための曖昧性問題の導入した日本語解析
 00 6(0.69):92-82-10 曖昧性問題のための曖昧性問題の導入した日本語解析
 00 7(0.72):92-87-01 曖昧性問題のための曖昧性問題の導入した日本語解析
 00 8(0.67):90-75-01 曖昧性問題のための曖昧性問題の導入した日本語解析
 00 9(0.67):90-75-01 曖昧性問題のための曖昧性問題の導入した日本語解析
 00 10(0.63):92-87-02 曖昧性問題のための曖昧性問題の導入した日本語解析

↑

ただし、↑a: a位階級上、↑n: n位階級下、-0: 階級化なし

B:例7 ↓ 新近向上文書の優先表示(↑3以上の表示を省略する)

↓

04 1(0.95):90-75-01 Discrimination network上での音韻的曖昧性問題について
 03 2(0.93):92-81-08 曖昧性を用いた曖昧性問題
 02 3(0.77):90-80-04 文脈を用いた曖昧性問題の「アプローチ」
 01 4(0.74):81-82-11 アスペクトをもとにした文脈解析
 00 5(0.73):81-84-02 曖昧性問題のための曖昧性問題の導入した日本語解析
 00 6(0.69):92-82-10 曖昧性問題のための曖昧性問題の導入した日本語解析
 00 7(0.72):92-87-01 曖昧性問題のための曖昧性問題の導入した日本語解析
 00 8(0.67):90-75-01 曖昧性問題のための曖昧性問題の導入した日本語解析
 00 9(0.67):90-75-01 曖昧性問題のための曖昧性問題の導入した日本語解析
 00 10(0.63):92-87-02 曖昧性問題のための曖昧性問題の導入した日本語解析

↑

(順位)(曖昧性)(文書ID) (文書タイトル)

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☒ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☒ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.